Weekly Report

November 5, 2017

1 Work

This week, I try to modify the iterative formula of bhsne (a faster version of tsne). bhsne speeds up the calculation of p_{ij} by knn graph (from $O(N^2)$ to $O(kN)$) and speeds up the calculation of q_{ij} by quadtree (from $O(N^2)$ to $O(NlogN)$). I want to compute q_{ij} by knn graph to reduce the complexity which means ignoring the q_{ij} between too far points. However, the result shows that q_{ij} between points with a long distance is an important factor to separate them Thus, I randomly select some points to compute q_{ij} instead of computing the distance between all points.

Also, I have a discussion with 付聪. We plan to speed up the knn graph construction with Product Quantization and accelerate the embedding process by push a group of close points (kmeans) in each iteration.

1.1 工作进度

Table 1: 工作进度

TASK	PROGRESS	DATE
dimension reduction	combine product quantization	11.30
hihgway project		12.10
*2Vec survey	collect papers	12.30

2 Paper Reading

2.1 Sound-Word2Vec: Learning Word Representations Grounded in Sounds

文章提出的方法学习了单词及其相关拟声词之间的关系，这是基于对同一段声音的不同标签学习了一个相似性。

word	word2vec	sound-word2vec
apple	apples, pear, fruit berry, pears, strawberry	bite, snack, chips chew, munch, carton
wood	lumber, timber, softwoods, hardwoods, cedar, birch	wooden, snap, knock, smack, whack, snapping
bones	skull, femur, skeletons, thighbone, pelvis, molar	eggshell, carrot, arm blood, polystyrene, crunch
glass	hand-blown, glassware, tumbler, Plexiglass, wine-glass, bottle	shattered, ceramic, smash clink, beer, spoon

Figure 1: 1

2.2 Sentiment Analysis of Citations Using Word2vec

本文对于文献的引用学习了嵌入表达，基于这样的表达执行了一个测试（分类引文是提出新方法还是改进已有算法）。

2.3 Ngram2vec: Learning Improved Word Representations from Ngram Co-occurrence Statistics

word2vec是像一个单词转换为一个向量，但是我们在文字使用中往往会有几个连起来使用的短语，文章将单纯的word扩展为多个word组成的Ngram进行学习。

2.4 A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications

本文是关于图嵌入的综述，主要包括三个方面：数据（图的各种类型），嵌入方法（矩阵分解、深度学习等），应用（节点分类，边预测等）。

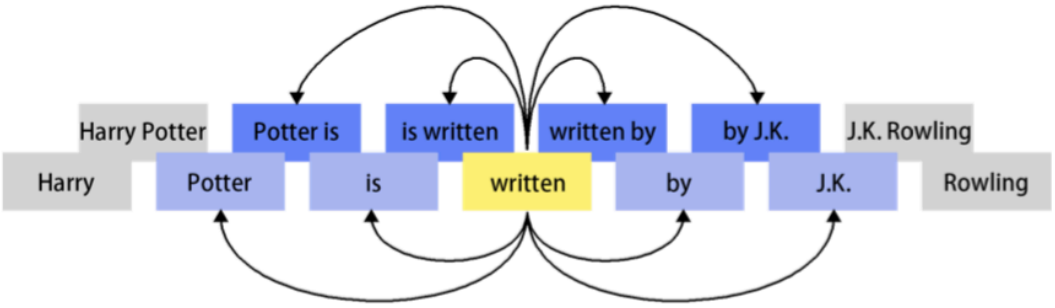

Figure 2: 3

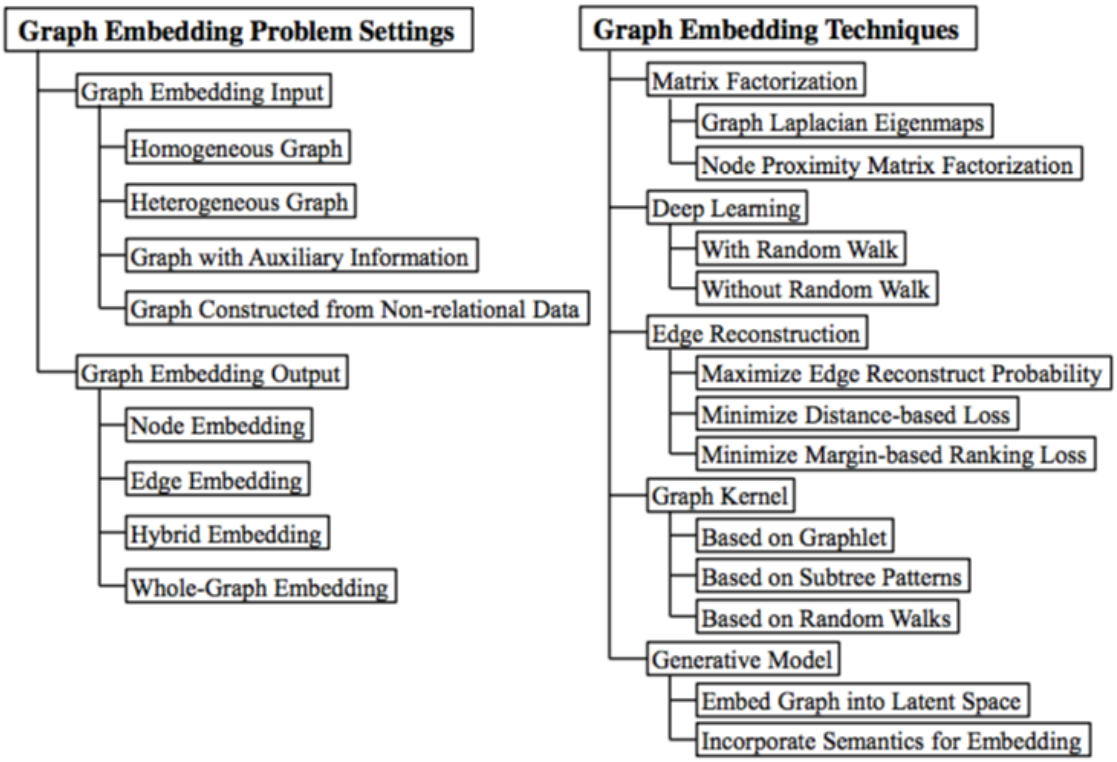

Figure 3: 4